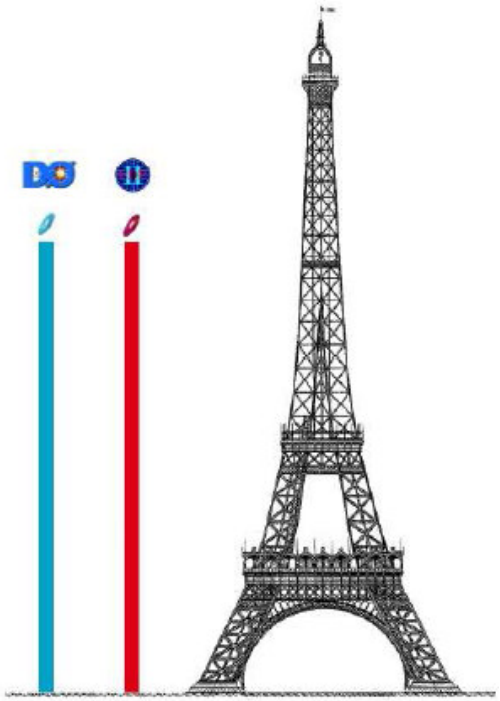
DØ Remote Computing

- Outline
  - Why ?
  - How ?
    - Tasks
    - Tools
    - When
  - Conclusion



(Part of an ongoing process)

Introduction - I

- The why

  Effective computing enables physics

- The Challenge
  - Large data volumes ~PB /yr (have 1000 M events)
  - Many cpus needed ~ 1.5M SpecInt (~3THz)

- The Complication
  - Computing distributed across 18 countries

- Current areas / concerns
  - Increased functionality
  - Expect x10 more data, at higher instantaneous L
  - Manpower and cpu moving to LHC → shared resources
  - So increasing use of common solutions / grid

Introduction – II – the context

- ~1997 – The Original Plan
  - All Monte Carlo to be produced off-site
  - SAM to be used for all data handling
- ~ 2002 – Offsite Analysis Task Force
  - Increased off-site computing – Regional Analysis Centres
    - Monte Carlo, data re-processing, data analysis
    - GridKa established as prototype
  - Run II 2002/3 (Bird) Reviews – strong praise
- Recent Run II (Shank) Review - Strong praise:
  - Use of SAM
  - Off-site reprocessing, using SAMGrid
  - Move towards common solns / being fully grid enabled

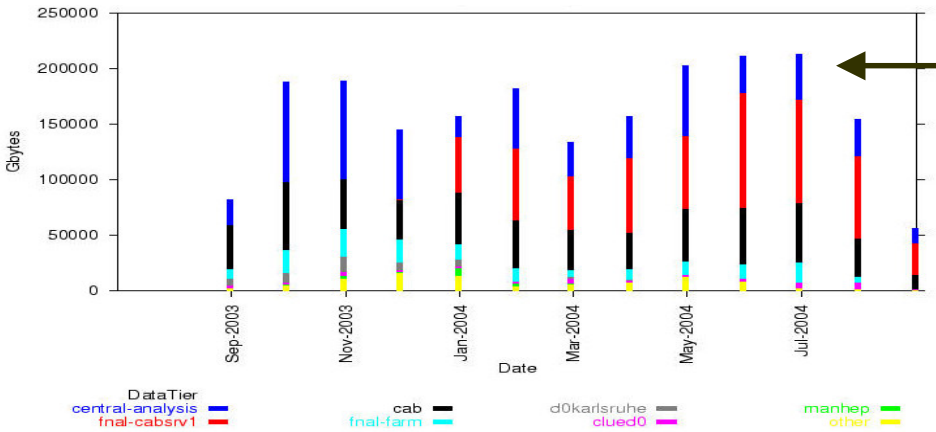
Must provide a production service, whilst improving functionality ✓

Introduction – III – the "tools"

- SAM (Sequential Access to Metadata)
  - Well developed metadata and distributed data replication system
  - Developed by DØ & FNAL-CD
- JIM (job & information management)
  - SAM + JIM → SAMGrid – computational grid
- Runjob
  - Handles job workflow management
- Additional tools
  - dØtools – User interface for job submission
  - dØrte – Specification of runtime needs

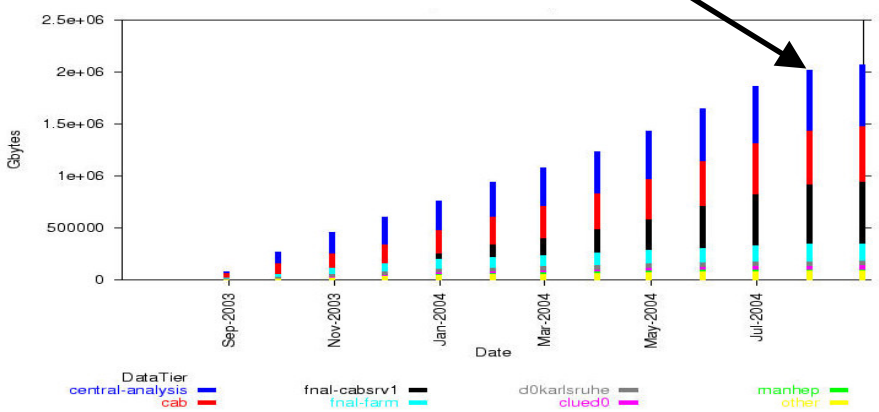
SAM Plots

40 active DØ SAM sites worldwide



Up to 200TB/month

Over 2 PB in last yr



Station	Snapshot Create Time	Requested Files	Projects (tot | run)	Projects Health (ok, error, waiting)	Last File Delivery	Deliveries
cab (history)	Wed Oct 20 04:06:26 2004	1000	10 | 4		Wed Oct 20 03:03:10 2004 (1h 3m) juste-19Oct2004164936	
central-analysis (history)	Wed Oct 20 04:07:03 2004	345	5 | 5		Wed Oct 20 04:06:28 2004 (35s) besancon_20041020022122	
central-router (history)	Wed Oct 20 04:07:33 2004	6	0 | 0	---	---	---
clued0 (history)	Wed Oct 20 04:06:11 2004	89	20 | 10		Wed Oct 20 04:05:15 2004 (56s) busato-pick_event-03-42-02-20Oct2004	
fnal-cabsrv1 (history)	Wed Oct 20 04:07:51 2004	104	67 | 55		Wed Oct 20 04:05:12 2004 (2m 39s) ajenkins-20Oct2004035616	
fnal-farm (history)	Wed Oct 20 04:10:44 2004	50	6 | 3		Tue Oct 19 19:41:06 2004 (8h 29m) Reco_18191.1098216771	
rp-router (history)	Wed Oct 20 04:07:43 2004	0	0 | 0	---	---	---

SAM TV

Monitors SAM & SAM stations worldwide

Monitored by (remote) SAM shifters

2nd generation on its way

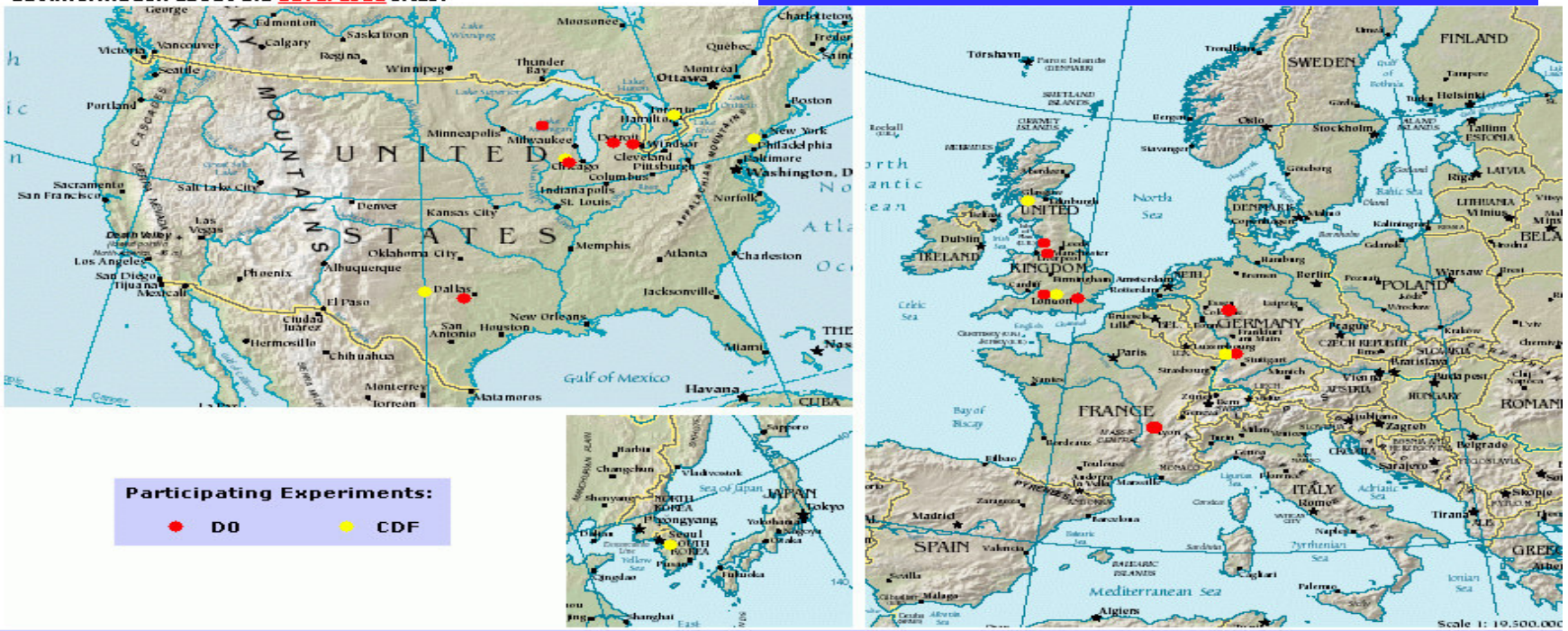
SAMGrid plots

SAM GRID INFORMATION & MONITORING SYSTEM

Launching the Monitoring System:

Please click at the map to monitor the execution sites.
Get information about the submission sites
Get information about the advertised sites.

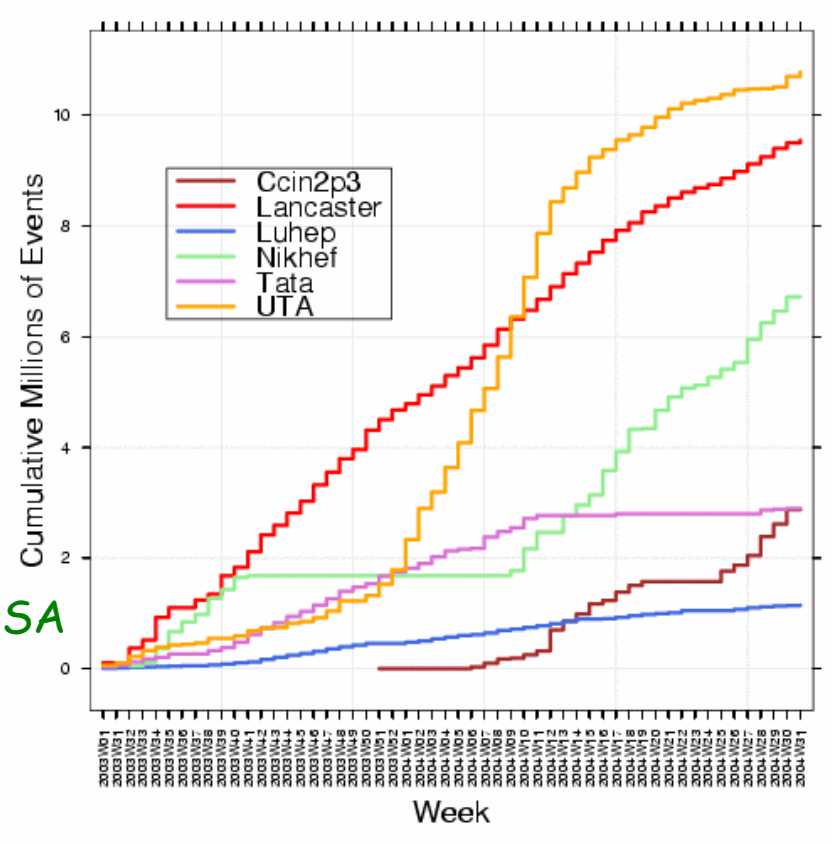
JIM: Active execution sites: ~12DØ



http://samgrid.fnal.gov:8080/

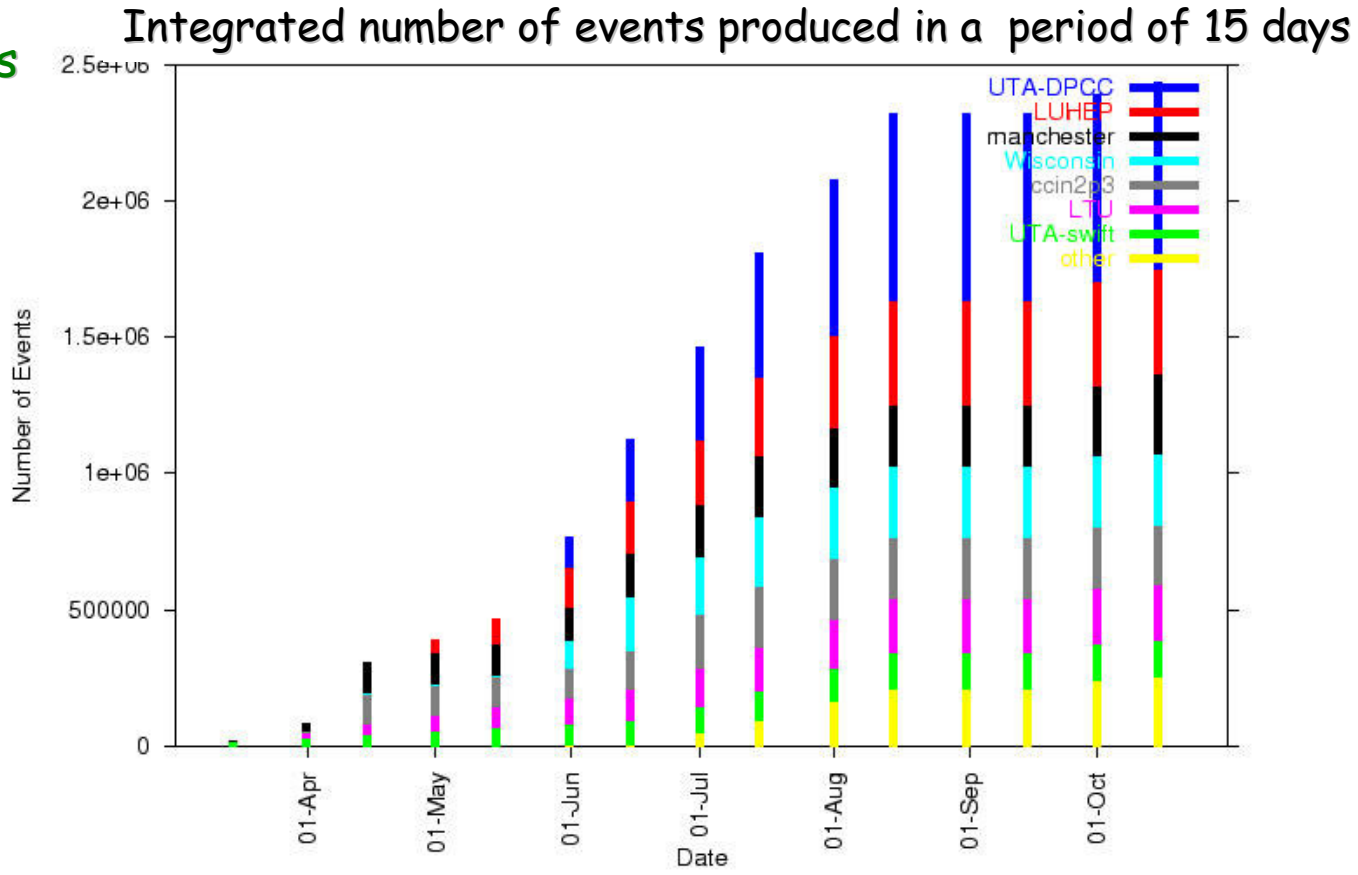
SAMGrid / Monte Carlo

- All MC produced off-site
  - Based on request system and jobmanager - mc_runjob
  - MC software package retrieved via SAM
  - Used 13 sites, in 8 countries last yr
  - Expect 4 new sites in next yr
- 37 M evts last yr (~2.5M with JIM)
- 1.0 - 1.5 M evts/week
- JIM now default at many MC centres
  - Running at >10 sites, inc Cz, D, Fr, UK, USA
    - more on way, inc central farm
  - Production efficiency ~90%
  - Grid infrastructure losses ~1-5%



SAMGrid Monte Carlo

2.5 M events



- For more details, see
  - http://www-d0.fnal.gov/computing/grid/deployment-issues.html

Reprocessing

- P14 Autumn 2003

P14 Reprocessing Status as of 26-Apr-2004 (Remote sites only)

Processed Events	97619114	
Sites	fnal ccin2p3 gridka nikhef uk westgrid	

  - 500M events total, 100M remote (actually processed in < 2months)
  - Based around mc_runjob
  - Done from DST
  - Distributed computing rather than Grid

- P17 winter 2004/5 - p17 released and under test
  - Aim for early 05 start, lasting ~ 6 months
  - x 10 larger, full 1000M events = 250TB remotely
  - Do from raw, need db proxy servers
  - SAMGrid as default
  - Use shared (LCG) resources

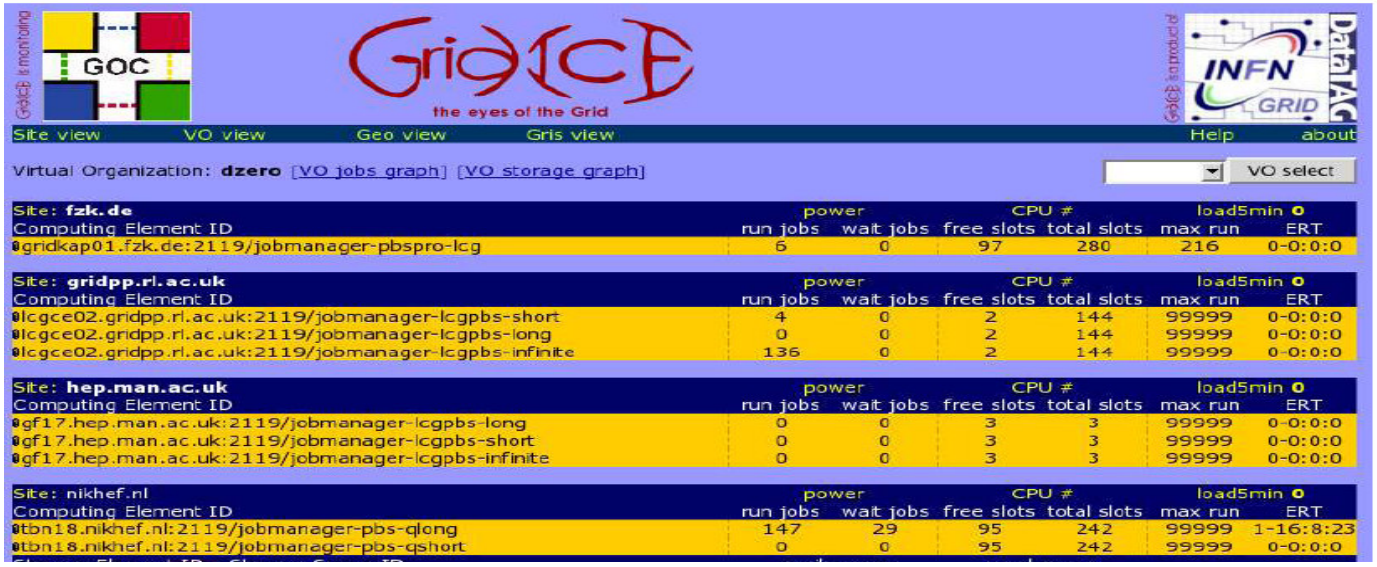
Massive Undertaking !

DØ – Production - LCG

- Increasing effort to ensure SAMGrid interoperability
  - Particularly with LCG, but also other major grids
  - MC generated on EDG/LCG and other shared resources "by hand"
  - Demo of sam_client functionality on LCG at London workshop in Apr
  - Will use LCG resources for p17 data reprocessing (as did for p14)

All Nikhef MC produced this way

LCG DZero VO



Common solutions / evolution

- Runjob:
  - mc_runjob currently used by SAMGrid for MC and reprocessing
  - Now a joint CDF?, CMS, DØ, FNAL-CD runjob project
  - dØrunjob - the DØ specific rewrite
  - Base classes from common runjob package
  - Available this autumn

Common projects:
Necessary, but potential new risk
(reliance on other stakeholders)

- Integration of tools: runjob, dØtools, dØrte, SAMGrid
- Central farm to run with SAMGrid
- Central submission / monitoring – SAM shifters / global monitoring ?
- Important steps in our computing evolution
  - MC → Reprocessing → fixing → analysis (on grid) ?
  - Using common tools / grid / shared resources

Conclusions

- Remote computing on this scale challenging
  - Problems as much sociological as technical
  - Conflicting needs of a production experiment

- Increased functionality
- Larger data sets, higher inst. L
- Shared resources
- Limited, non-dedicated manpower

→ Common solutions / standard tools / use of the grid

Even greater challenge, but making good progress.

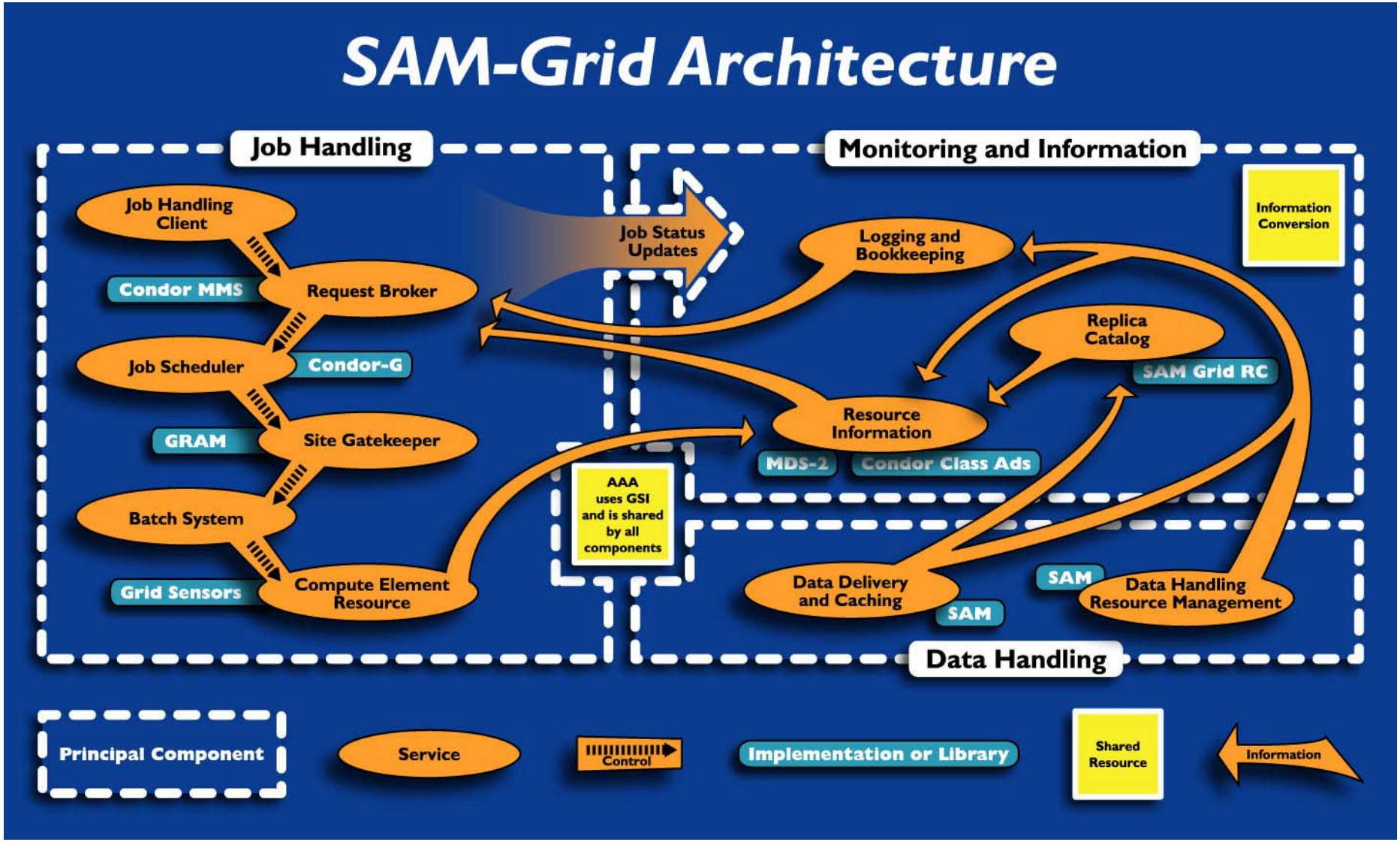
→ **More physics**

(+ DØ can continue to be a world force in grid computing)

Back up slides

In no particular order

SAMGrid Architecture



P14 (Round I) reprocessing lessons

- Planning for reprocessing run began June 2003
- Required ~5 ½ months preparation until startup
  - Most time spent on making p14 version of reco sufficiently robust to proceed
  - Significant changes in reco capabilities lead to rethinking entire processing chain and priorities
  - Began reprocessing ~15 Nov 2003, finished ~5 Jan 2004

    ~100M events processed at remote sites

    ~25TB data transferred

    Final merging, storage of TMBs done at FNAL to reduce load on remote sites, processing done from DSTs (no Db access)
  - ~ ½ FTE required at each remote site for duration of processing phase